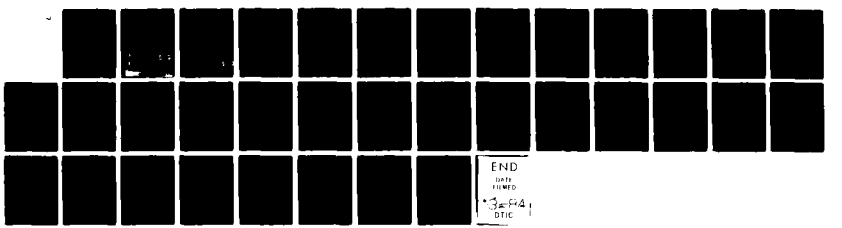
AD-A137 994 A COMPARISON OF RANDOM BALANCE AND TWO-STAGE GROUP SCREENING DESIGNS PART 2(U) DESMATICS INC STATE COLLEGE PA C A MAURO ET AL. FEB 84 TR-113-14 N00014-79-C-0650 1/1

UNCLASSIFIED F/G 12/1 NL

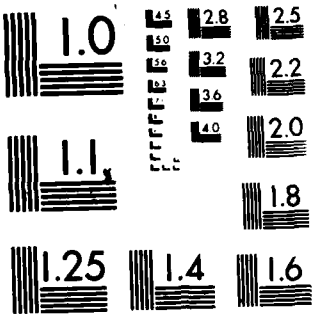


END

DATE FILMED

4-84

DTIC



(12)

A COMPARISON OF RANDOM BALANCE AND TWO-STAGE GROUP SCREENING DESIGNS - PART II

by

Carl A. Mauro
and
Kevin C. Burns

— STATISTICS —

— OPERATIONS RESEARCH —

— MATHEMATICS —

DESMATICS, INC.

P.O. Box 618
State College, PA 16801

DESMATICS, INC.

P. O. Box 618
State College, PA. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

A COMPARISON OF RANDOM BALANCE AND TWO-STAGE GROUP SCREENING DESIGNS - PART II

by

Carl A. Mauro
and
Kevin C. Burns

TECHNICAL REPORT NO. 113-14

February 1984

This study was supported by the Office of Naval Research under Contract No. N00014-79-C-0650, Task No. NR 042-467

TABLE OF CONTENTS

1. INTRODUCTION

Experimental studies involving a large number of factors can require a prohibitively large and costly research program. Often it is anticipated that only a small subset of the factors is important in explaining the response. Accordingly, we may want to conduct a preliminary screening experiment to determine the subset of "most important" factors. Such experiments are not an end in themselves but are performed as an initial phase of experimentation. Once the most influential factors have been isolated, future experimentation can then investigate these factors in detail. By reducing the size of the experimental problem at the screening stage, we are able to conserve resources and more efficiently and effectively study the factors of interest. Screening experiments have potential application in many research areas such as manufacturing, engineering, product development, and simulation.

The factor screening problem has been considered by a number of authors; see, for instance, Anscombe (1963), Booth and Cox (1962), Budne (1959b), Kleijnen (1975), and Satterthwaite (1959). However, there has been no objective evaluation and comparison of the available screening methods. Kleijnen (1975) and Smith and Mauro (1984) have divided the screening problem into two general situations. These are the unsaturated/saturated and supersaturated situations. In the unsaturated/saturated situation, one can afford to invest more runs than there are factors. Designs that have been generally recommended for use in this situation include Plackett-Burman designs (Plackett and Burman 1946) and resolution IV foldover designs (Box, Hunter, and Hunter 1978).

These designs have been extensively studied and their properties are well known.

In the supersaturated situation, the number of factors equals or exceeds the number of runs available for screening. Designs satisfying this limitation which have been suggested include group screening designs (Li 1962; Patel 1962; Watson 1961), random designs (Satterthwaite 1959), and systematic supersaturated designs (Booth and Cox 1962). The performance characteristics of these designs are largely unknown. Furthermore, there are few examples of supersaturated screening experiments in the literature. For the researcher contemplating a supersaturated experiment, it is therefore difficult to find practical guidelines for either design or analysis.

In this paper we compare the performance of random balance (RB) and two-stage group screening (GS) designs in a case study in which K=100 factors are screened in N=20,42,62, and 84 runs. In addition, we discuss the relative merits and demerits of each approach. This discussion should provide some practical insight into the selection and use of these designs. We have not included systematic supersaturated designs in our study since these designs have not been tabulated for K>36 and there is no efficient algorithm for their general construction. Booth and Cox (1962) have shown that these designs are in general more efficient than RB designs.

As an underlying screening model, we will assume the model defined in Section 2. In Sections 3 and 4 we describe and discuss the RB and GS strategies which we consider. In Section 5 we present and discuss the results of our case study. A brief summary follows in Section 6.

2. A SCREENING MODEL

It generally suffices in screening problems to employ two levels of each factor; see, for example, Montgomery (1979; p. 5) and Box, Hunter, and Hunter (1978; pp. 306-307). Furthermore, for the purpose of detecting the factors which have a major effect it is usually reasonable to assume the first-order model

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} + \varepsilon_i \qquad (2.1)$$

where y_i is the $i^{\underline{th}}$ observation, β_0 is a constant component common to all observations, K is the number of factors, β_j is the linear effect of the $j^{\underline{th}}$ factor, x_{ij} is the level (coded ±1) of the $j^{\underline{th}}$ factor in the $i^{\underline{th}}$ run, and the ε_i are i.i.d. $N(0,\sigma_\varepsilon^2)$ error terms, σ_ε^2 unknown. Ordinarily, we would use model (2.1) over a relatively small region of the factor space.

In this paper we restrict our comparisons of screening strategies to model (2.1). Moreover, we will use this model as a basis for performance assessment and data generation in our case study.

3. RANDOM BALANCE DESIGNS

In a two-level (±1) RB design, each column of the design matrix consists of N/2 +1's and N/2 -1's where N, an even number, denotes the total number of runs to be made. The +1's and -1's in each column are assigned randomly, making all possible combinations of N/2 +1's and N/2 -1's equally likely, with each column receiving an independent randomization.

The principal advantages of RB sampling for use in screening problems are its flexibility and the ease with which we can prepare these designs. The number of runs N can be selected independently of the number of factors K; no mathematical restriction or relationship (except that N be even) need exist between N and K.

There are two primary disadvantages to RB sampling. The first of these is that confounding is random. Anscombe (1959; p. 201) has written, "The fact that the degree of nonorthogonality or unbalance is random can be made the basis for an objection to the whole notion of random balance designs. Such designs may work well on the average, but should I trust to one on this occasion?" The second disadvantage, which is closely related to the first, is that there is no generally accepted method of analysis for RB designs. The simplest approach, and the one adopted in this paper, is to consider each factor separately, ignoring all other factors, and apply a standard F-test. More sophisticated analysis methods which have been used include least squares stepwise and stagewise regression. We refer the reader to Youden, et.al. (1959) for a more complete discussion of RB experimentation.

As we have already indicated, we consider a standard F-test applied separately to each factor as the method of analysis for RB designs. We assume for simplicity that each F-test is conducted at the same level of significance, say α_r. We denote such a strategy by RB(N, α_r). Furthermore, for screening purposes, we classify a factor as important only if its associated F-ratio is significant, i.e., equals or exceeds the upper $100(1-\alpha_r)$ percentage point of an F-distribution with (1,N-2) degrees of freedom.

The simple least squares estimator of β_j obtained by ignoring all other factors is given by

$$\hat{\beta}_j = (\bar{y}_{+j} - \bar{y}_{-j})/2 \tag{3.1}$$

where $\bar{y}_{+j}(\bar{y}_{-j})$ is the average value of the response over the N/2 runs at the +1(-1) level of the $j^{\underline{th}}$ factor. To simplify notation, we let $\underline{y}$ denote the Nx1 vector $(y_1,y_2,\ldots,y_N)'$ of responses and $\underline{x}_j$ denote the Nx1 vector $(x_{1j},x_{2j},\ldots,x_{Nj})'$. In an RB experiment, the NxK design matrix $\underline{X}=[\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_K]$ is, by construction, stochastic. Assuming that $\underline{X}$ and the ε_i are independent, it is easily shown that conditional on $\underline{X}$,

$$E(\hat{\beta}_j|\underline{X}) = \beta_j + \Big(\sum_{i\neq j} \beta_i \underline{x}_i'\underline{x}_j\Big)/N \tag{3.2}$$

and

$$V(\hat{\beta}_j|\underline{X}) = \sigma_\varepsilon^2/N. \tag{3.3}$$

The conditional mean square error (MSE) of $\hat{\beta}_j$ is then

$$MSE(\hat{\beta}_j | \underline{X}) = \sigma_\epsilon^2/N + \Big(\sum_{i \neq j} \beta_i \underline{x}_i' \underline{x}_j \Big)^2 / N^2 \qquad (3.4)$$

Unconditionally, Mauro and Smith (1984) have shown, see also Box (1959), that

$$E(\hat{\beta}_j) = \beta_j, \qquad (3.5)$$

and

$$V(\hat{\beta}_j) = \tau_j^2/(N-1) + \sigma_\epsilon^2/N \qquad (3.6)$$

where $\tau_j^2 = \sum_{m \neq j}^{K} \beta_m^2$.

As Box (1959) points out, equations (3.2) and (3.3) refer to the behavior of the estimates for repetitions of a particular RB design. Equations (3.5) and (3.6), on the other hand, refer to the behavior of the estimates if we average over the random choice of RB designs. Box (1959) comments further that although $\hat{\beta}_j$ is unconditionally unbiased, the effect of the conditional bias term in (3.2) is transferred to the (unconditional) variance of $\hat{\beta}_j$ which now contains terms from every other factor present.

4. TWO-STAGE GROUP SCREENING DESIGNS

In the two-stage group screening method, introduced by Watson (1961), the individual factors (each at two levels) are partitioned into groups. By assigning the same level to all component factors, the groups are tested (in a first-stage experiment) as if they were single factors. Those factors within significant groups are subsequently tested individually in a follow-up second-stage experiment. The basic idea behind this method is that factors within a group are completely confounded. Thus, after the first stage, we can eliminate from further consideration all factors within non-significant groups. The fewer the important factors, the more effective is the technique.

To study the group-factors in the first stage and the individual factors which reach the second stage, we will use the resolution III multifactorial designs of Plackett and Burman (PB). These designs are specially constructed two-level orthogonal designs for studying up to (4m-1) factors in 4m runs. Mathematically, the number of runs required by the smallest PB design to study S factors (or group-factors) is given by

$$B(S) = S+4 - S(\text{mod } 4).$$

However, in order to insure at least one degree of freedom for error, we shall employ the PB design in B(S+1) runs. Since we are assuming an underlying first-order screening model, the use of PB designs would seem reasonable. We analyze these designs with the usual analysis of variance procedures for factorial experiments.

We make the additional assumption that the K factors are partitioned randomly into G groups of size g; if K is not a multiple of g, we will assume that the group sizes are taken as "evenly" as possible. The assumptions of equal group sizes and random allocation to groups are appropriate if there is no prior knowledge about the effects. Of course, if the experimenter has prior information indicating that some effects are larger than others, then it is better to group those effects together. Moreover, the experimenter is not limited in practice to a constant group size. Watson (1961) has discussed the device of using different group sizes when prior probabilities differ.

An important advantage of GS designs is that we can to some extent control the confounding pattern. There are two disadvantages to GS designs. The first of these is that the total number of runs required in a group screening experiment is random. Specifically, the number of second-stage runs depends on the number of significant groups, which is subject to testing and grouping variation. Thus, unlike in an RB design, the total number of runs required in a GS design is not fixed prior to experimentation. The second disadvantage is that effects may cancel within a group. As a simple example, consider two factors which have effects that are negatives or near-negatives of each other. If these two factors are the only important factors in a given group, their effects will cancel or their combined effect may be masked by experimental error. It is desirable, then, to have prior knowledge of the directions of all suspected effects. With this information, factor levels can be assigned so that all potential effects are in the same direction.

In practice, the direction will most likely be known for some suspected effects and unknown for others. Kleijnen (1975; p. 489) has pointed out that "unequal groups sizes make it possible to test a factor individually when we do not know the direction of its effect." However, it would seem feasible to treat only a few suspected effects in this manner. It seems certain that in some applications cancellation may not be avoided. The effect of cancellation has been studied under "worst-case" assumptions by Mauro and Smith (1982) and Mauro (1983a).

Finally, we let α_1 and α_2 denote the levels of the significance tests performed at the end of the first and second stages, respectively. Since we assume a constant group size g in addition to random allocation, our GS strategy is completely specified by g, α_1, and α_2. We denote such a strategy by GS(g, α_1, α_2).

5. A CASE STUDY

This section describes a case study in which we compare RB(N,α_r) strategies for testing K=100 factors in N=20,42,62, and 84 runs with corresponding GS(g, α_1, α_2) strategies in as many expected runs. In addition, for each N, we consider type I error rates of 5% and 10%. Power curves are used as a basis of comparison.

In our case study, we assume model (2.1) together with the regression coefficients given in Table 1. The absolute magnitudes of the effects listed in Table 1 correspond to the expected order statistics (rounded to two decimal places) of a sample of size 100 from a gamma distribution with mean $.5\sigma_\varepsilon$ and standard deviation $1.5\sigma_\varepsilon$, see Figure 1. These particular effects seem reasonable as an illustrative screening example and are in accordance with the "mal-distribution" assumption discussed by Budne (1959a). Furthermore, since it is unlikely that all effect directions will be known a priori, we have allowed some factors to have negative effects. The proportion of negative effects has been made a decreasing function of their absolute magnitudes, since that scenario seems most likely to occur in practice.

5.1 RB Power Curves

In order to evaluate the power of an RB(N,α_r) strategy, we adopt the method of averaging over the set of possible RB(N,α_r) designs. (Dempster 1960 has discussed the ramifications of this approach.) In the appendix, we show that the RB model can be related to the exchangeable

No.	β/σ_ε	No.	β/σ_ε	No.	β/σ_ε
48	0.00	1	0.14	1	-0.82
4	-0.01	1	-0.15	1	0.92
2	0.01	1	0.17	1	-1.04
1	-0.02	1	-0.20	1	1.18
3	0.02	1	-0.22	1	-1.34
2	-0.03	1	-0.25	1	1.52
2	0.04	1	-0.28	1	1.74
1	-0.05	1	0.32	1	-2.00
1	0.05	1	-0.36	1	2.31
1	0.06	1	0.40	1	2.71
1	-0.07	1	-0.45	1	3.20
1	-0.08	1	-0.51	1	-3.85
1	0.09	1	0.57	1	4.79
1	0.11	1	-0.64	1	6.28
1	0.12	1	0.72	1	9.57

Table 1. Regression Coefficients Used in Case Study.

linear model, as discussed by Arnold (1979, 1981). Using results derived for the exchangeable linear model, we obtain expressions which can be used to determine power probabilities in an RB(N,α_r) strategy. The power approximations outlined in the appendix are an improvement over those given in Mauro and Smith (1984). It is our experience that these approximations are very good even for small values of N. We make the additional observation that the type I error rate in an RB(N,α_r) strategy is very closely approximated by α_r.

Figures 2a-2d contain the power curves corresponding to the eight RB(N,α_r) strategies specified by N=20,42,62, and 84 runs with α_r=0.05 and 0.10. On each of these plots, the solid curve corresponds to α_r=0.05 and the dashed curve corresponds to α_r=0.10.

5.2 GS Power Curves

Mauro (1983a) has developed formulas to determine power probabilities in a GS(g, α_1, α_2) strategy for the special case in which each individual factor has an effect of size $+\Delta\sigma_\varepsilon$, $-\Delta\sigma_\varepsilon$, or 0. Unfortunately, the more general case of arbitrary effects is too complex to analyze mathematically. For our case study, therefore, it was necessary that we use a Monte Carlo simulation program to estimate power probabilities in a GS(g, α_1, α_2) strategy. Using this program we determined, by trial and error, the particular GS(g, α_1, α_2) strategies which maximized power subject to the specified constraints on the expected number of runs, E(N), and type I error.

The "optimal" GS(g, α_1, α_2) strategies that we selected are re-

ported in Table 2. In this table we also have given the sample mean $\bar{N}$ and standard deviation S_N associated with each GS strategy. The results are based on 10,000 simulations (except for one case noted in the table).

It is important to note that in a GS(g, α_1, α_2) strategy, the expected number of runs depends only on g and α_1. Furthermore, given g and α_1, type I error can be expressed as

$$\text{Type I Error} = C(g,\alpha_1)*\alpha_2.$$

Thus, type I error is directly proportional to α_2 and bounded by the constant $C(g,\alpha_1)$.

Figures 3a-3d contain the empirical power curves corresponding to the selected GS(g, α_1, α_2) strategies with 5% type I error. In these plots, the solid curve is the power associated with the positive effects. Since the power associated with a negative effect is generally less than that for a positive effect of the same magnitude, we have marked the negative effects separately. This phenomenon is due to cancellation and will be discussed later.

We do not present the GS power curves for 10% type I error since these curves were virtually identical, except for effects of small magnitude (less than $.5\sigma_\varepsilon$), with the curves plotted in Figure 3. Apparently, the levels of α_2 in the GS(g, α_1, α_2) strategies reported in Table 2 were sufficiently large to permit detection of even moderately sized effects for both 5% and 10% type I errors.

Intended Expected Number of Runs	0.05	0.10	$\bar{N}$	S_N
20	GS(20,0.00023,0.6150)	GS(20,0.00023,1.0000)	20.6	17.0
42	GS(6,0.00150,0.3125)	GS(6,0.00150,0.6250)	42.3	15.2
62	GS(5,0.00200,0.1667)	GS(5,0.00200,0.3333)	62.9	12.9
84	GS(4,0.05000,0.1135)	GS(4,0.05000,0.2270)	83.9*	14.6*

*Simulation program timed out after 5,736 simulations

Table 2. GS(g, α_1, α_2) Strategies Determined By Trial-and-Error Search

5.3 Discussion

It is convenient to define relative testing cost (RTC) as the ratio of the number (or expected number) of runs required by a screening strategy to B(K+1), which is the number of runs required by a PB design for K factors. A quick calculation will show that for K=100 factors, run numbers of 20,42,62, and 84 correspond roughly to relative testing costs of 20%, 40%, 60%, and 80%, respectively.

A comparison of Figures 2a-2d with 3a-3d shows that the GS power curves are clearly superior to the corresponding RB power curves for 60% and 80% RTC. At 40% RTC, the GS and RB power curves are fairly comparable, although the RB strategy has slightly greater power for large effects while the GS strategy has more power to detect small effects. For 5% type I error, the curves cross at about $4.3\sigma_\varepsilon$, at which power is roughly 50%. A comparison of the 20% RTC power curves shows that RB is the superior method here. In this case, the selected GS strategy has only a 25% chance to detect the largest effect ($9.57\sigma_\varepsilon$) in the model. On the other hand, the RB(20,0.05) strategy has roughly a 25% chance to detect an effect of size $4.5\sigma_\varepsilon$, a 50% chance to detect an effect of size $6\sigma_\varepsilon$, and a 95% chance to detect the largest effect.

As indicated by the uniformly low power attained at 20% RTC, group screening designs are not well suited for use when there are severe limitations on the number of runs. To screen 100 factors in 20 runs, for example, we note that group sizes less than seven cannot be used since the number of first stage runs must necessarily exceed 20 runs. For group sizes of seven or larger, the total number of runs required

by both stages of screening will exceed 20 if even one group is carried over to the second stage. Consequently, in order to have $E(N) \leq 20$ runs, α_1 must be extremely small, ensuring that the expected number of significant groups is less than one. This results in very low power, no matter how large an effect might be.

As mentioned previously, we have purposely included less negative effects in our case study than positive effects, with the proportion of negative effects being a decreasing function of magnitude. For very small effects, this proportion is nearly 50%, and the GS power curves show that small positive and negative effects have nearly the same probability of being detected. For larger effects, however, a smaller proportion of factor effects are negative. Thus, negative effects have a greater chance of being grouped together with positive effects, resulting in cancellation. Positive effects, on the other hand, are more likely to be grouped together, precluding cancellation. As a result, negative effects are detected less frequently than positive effects of the same magnitude. However, as can be seen from Figure 3, the effect of cancellation is minimal in this case study.

An important practical consideration which we have not yet addressed is that our determination of the "optimal" $GS(g, \alpha_1, \alpha_2)$ strategies listed in Table 2 required prior knowledge of the effects. Of course, if one's prior knowledge is perfect, there is no need for a screening experiment. More realistically, though, one's prior knowledge is never perfect and some speculation is required. In any event, it is hard to see how one might reasonably go about choosing a $GS(g, \alpha_1, \alpha_2)$ strategy in the absence of such prior information or speculation.

To indicate the potential effects of imprecise prior knowledge, consider a GS(g, α_1, α_2) strategy chosen with a specific distribution of effects in mind. If this set of effects does not closely approximate the true situation, the chosen group size may not be optimum. Furthermore, α_1 and α_2 may be misspecified, so that RTC and type I error deviate from their desired values. We refer to Mauro (1983b) for a study of this problem in the special case where each individual factor has an effect of $-\Delta\sigma_\varepsilon$, $\Delta\sigma_\varepsilon$, or 0.

An additional consideration in the use of a GS strategy is that the number of test runs is a random variable. In this paper we have restricted ourselves, somewhat arbitrarily, to looking at the expected number of runs. From a practical standpoint, however, this can be rather disconcerting to the researcher contemplating a group screening experiment. In order to evaluate the severity of this problem, we have estimated the standard deviation of the number of runs for each of the four cases considered. These quantities (S_N) are given in Table 2 and can be seen to be very large compared with RTC. In Figure 4 we present a histogram of the number of runs needed in each of the 10,000 simulations of the GS(5, 0.002, 0.1667) strategy (60% RTC and 5% type I error). Inspection of this histogram shows that in 10% of the simulations the number of runs was 84 or greater.

Since an experimenter might be reluctant to use a strategy which does not allow him to predetermine his testing cost, we have considered a modified version of group screening, where the number of runs is fixed. The experimenter decides beforehand the number of groups, say m, he is willing to carry over to the second stage. After the first-

stage experiment, the m groups with the largest estimated effects are chosen and their component factors tested individually in a second-stage experiment. This of course implies a random first-stage significance level. However, the overall type I error can be controlled by suitably adjusting the second-stage significance level. We have done some preliminary investigation of this type of strategy and our results indicate that its performance is comparable to that of standard group screening. However, further work needs to be done, including the investigation of hybrid strategies where only the maximum number of runs is specified, allowing a smaller experiment if such seems justified by the first-stage results.

6. SUMMARY

In this paper we have focused on the problem of supersaturated screening experiments. We have restricted our ttention to two basic screening methods: random balance and two-stage group screening. Our primary observations are illustrated by means of a case study in which K=100 factors are screened in N=20,42,62, and 84 runs. A comparison of power curves showed that group screening had much greater power in N=62 and 84 runs. Random balance had slightly greater power for detecting large effects in N=42 runs. When N=20 runs, group screening had only a 25% chance of detecting the largest effect ($9.57\sigma_\varepsilon$) in the model while random balance had a 95% chance of detecting that effect.

In addition to a comparison of power, we discuss the relative merits and demerits of each strategy. Our findings indicate that unless there is a severe limitation on the number of runs group screening appears to be the better strategy. However, there are two primary drawbacks to the use of group screening. First, the total number of runs is a random variable. Second, one needs a certain amount of prior information to choose an appropriate group screening strategy.

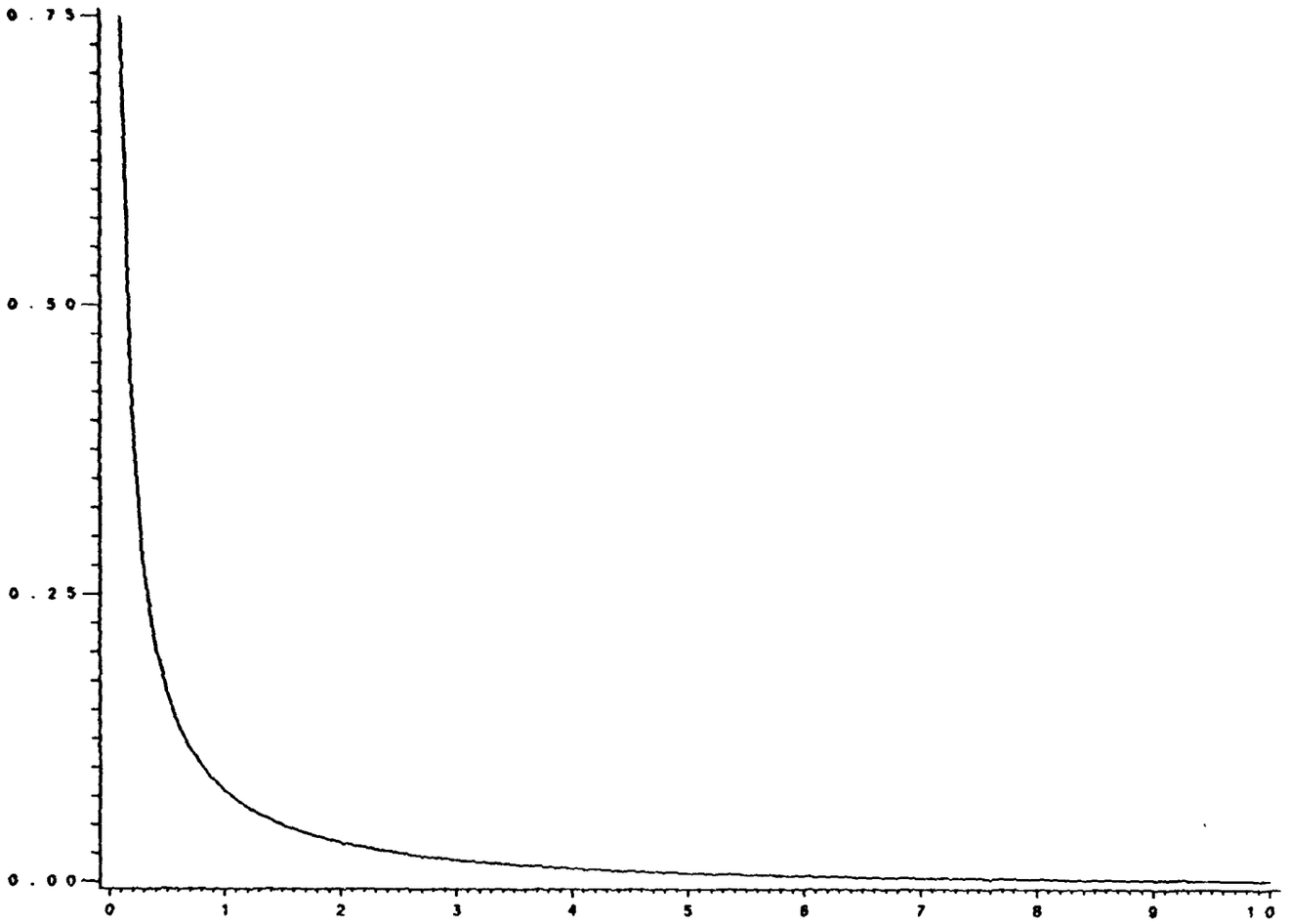
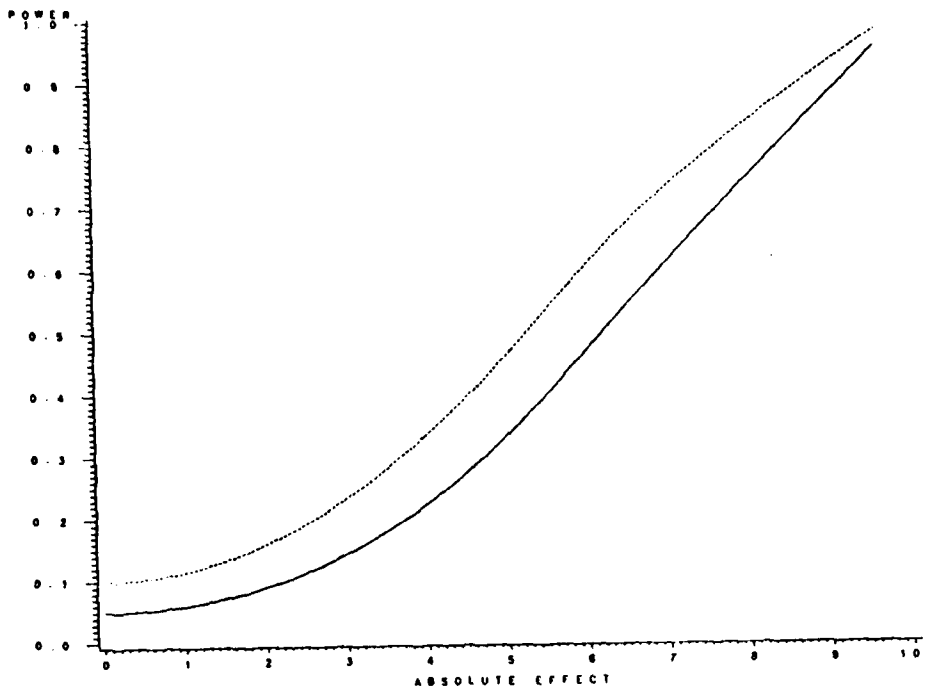
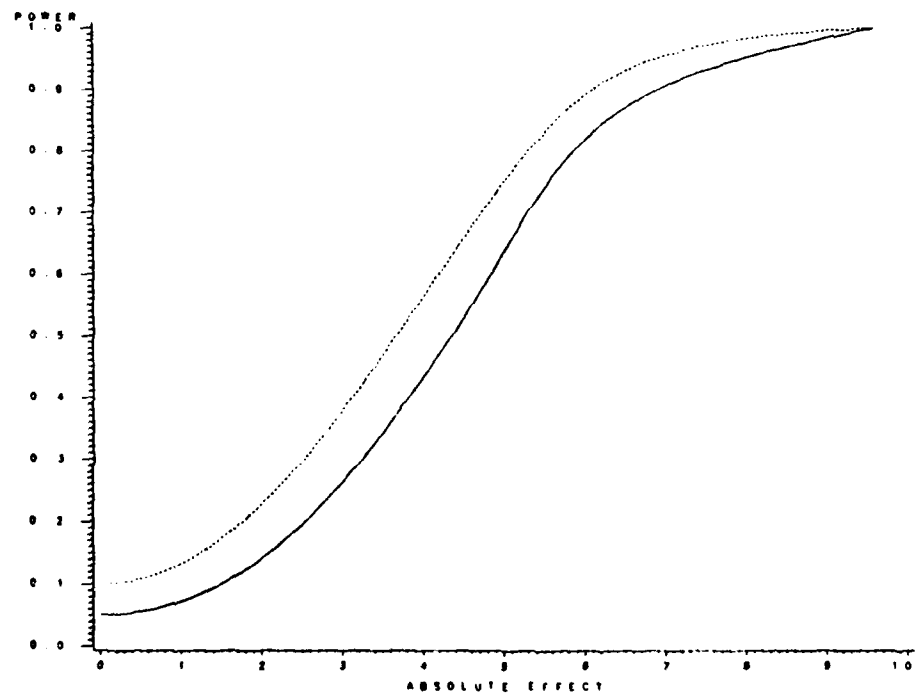


Figure 1. Probability Density Function of a Gamma Distribution with Mean $.5\sigma_\epsilon$ and Variance $2.25\sigma_\epsilon^2$.

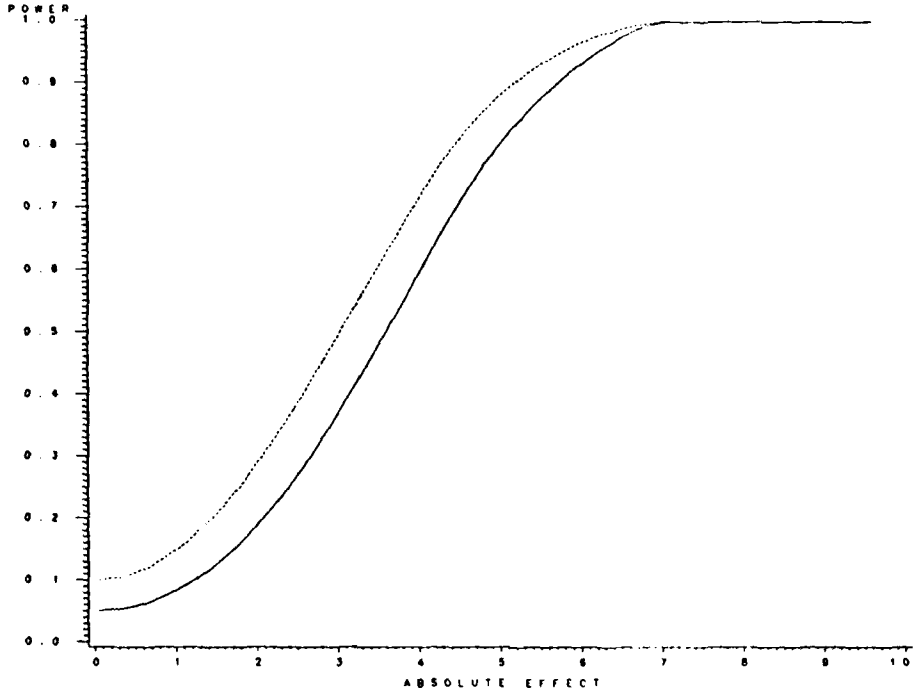


(a) N=20

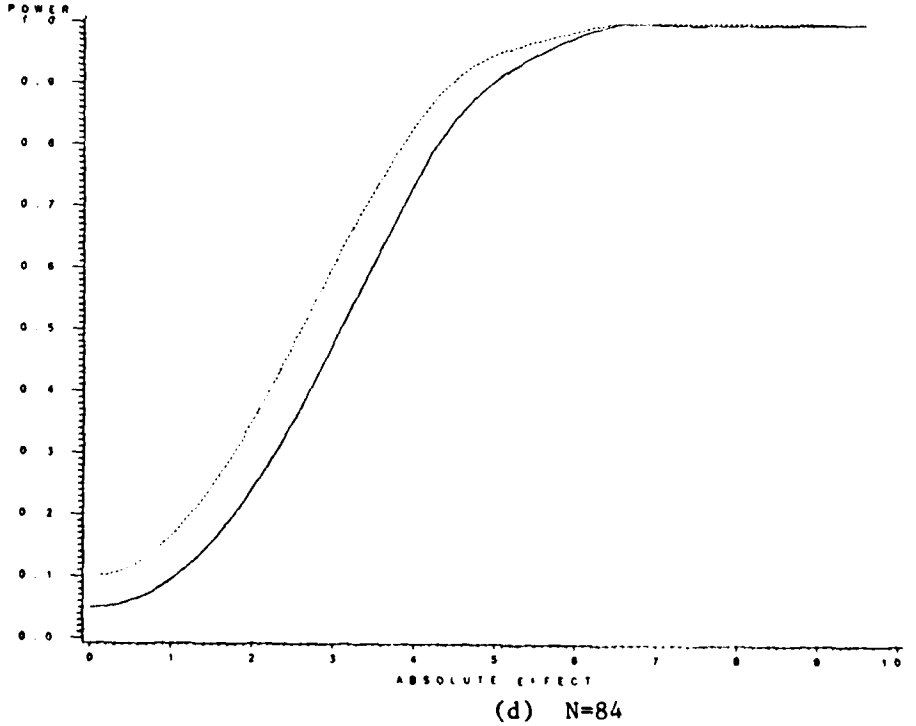


(b) N=42

Figure 2. RB Power Curves: Solid Curve Is For 5% Type I Error, Dashed Curve Is For 10% Type I Error.

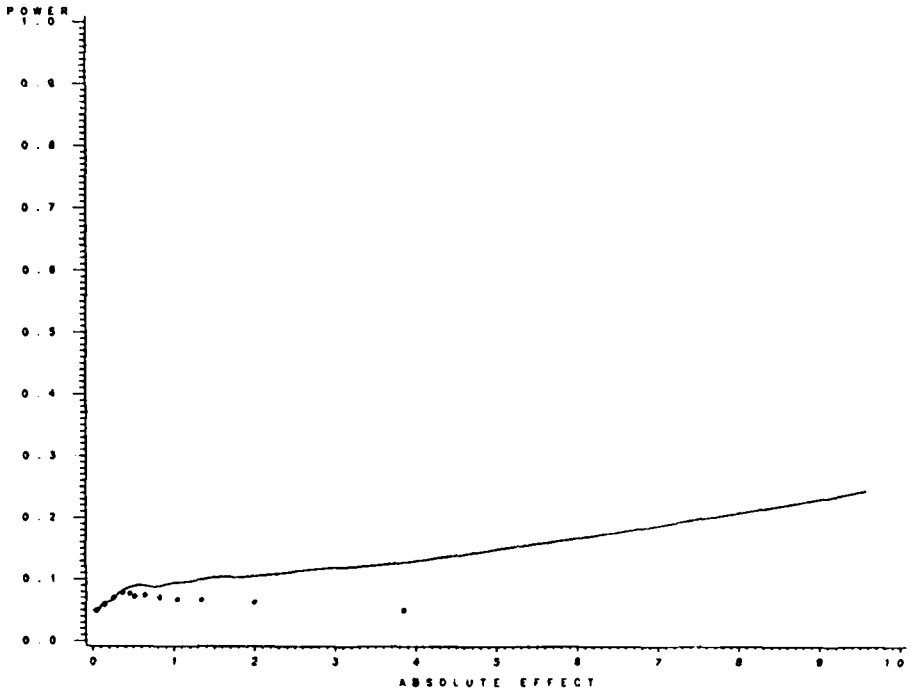


(c) N=62

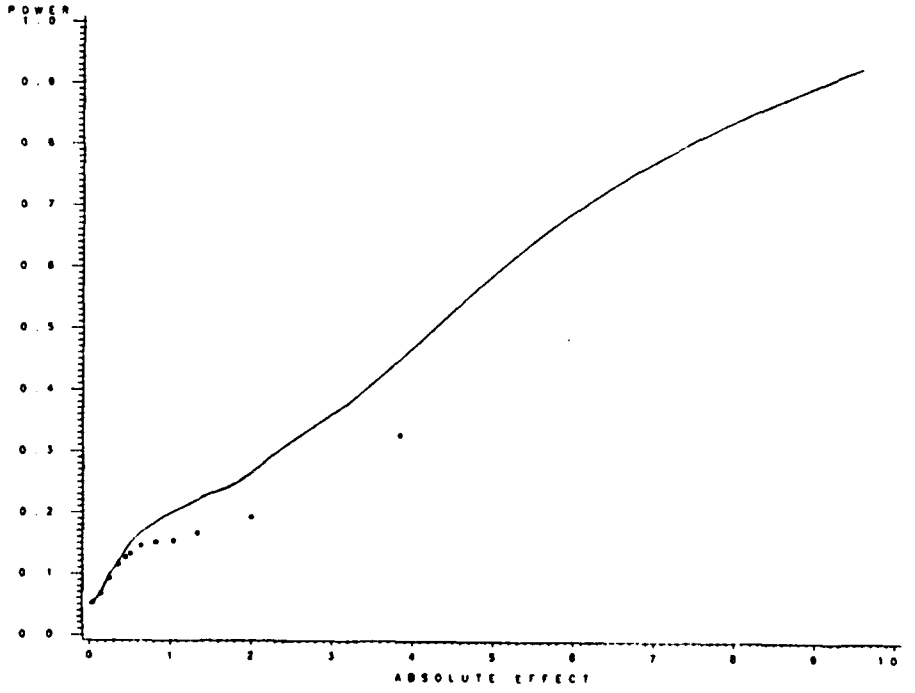


(d) N=84

Figure 2 (continued). RB Power Curves: Solid Curve Is For 5% Type I Error, Dashed Curve Is For 10% Type I Error.

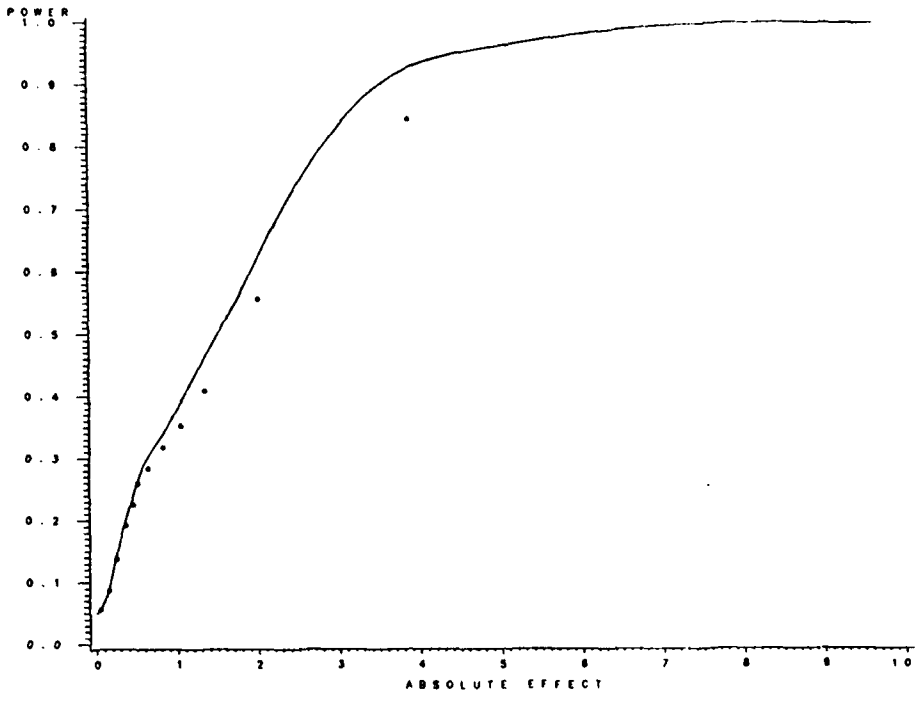


(a) E(N)=20

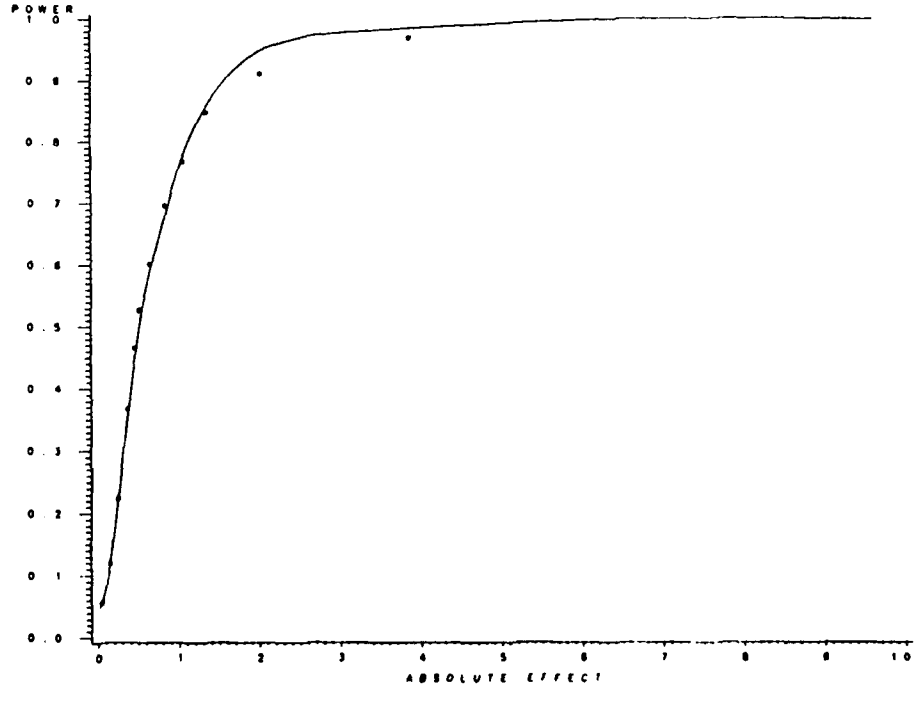


(b) E(N)=42

Figure 3. GS Power Curves for 5% Type I Error: Solid Curve Is For Positive Effects, Dots Are For Negative Effects.



(c) E(N)=62



(d) E(N)=84

Figure 3 (continued). GS Power Curves For 5% Type I Error. Solid Curves Are For Positive Effects, Dots Are For Negative Effects.

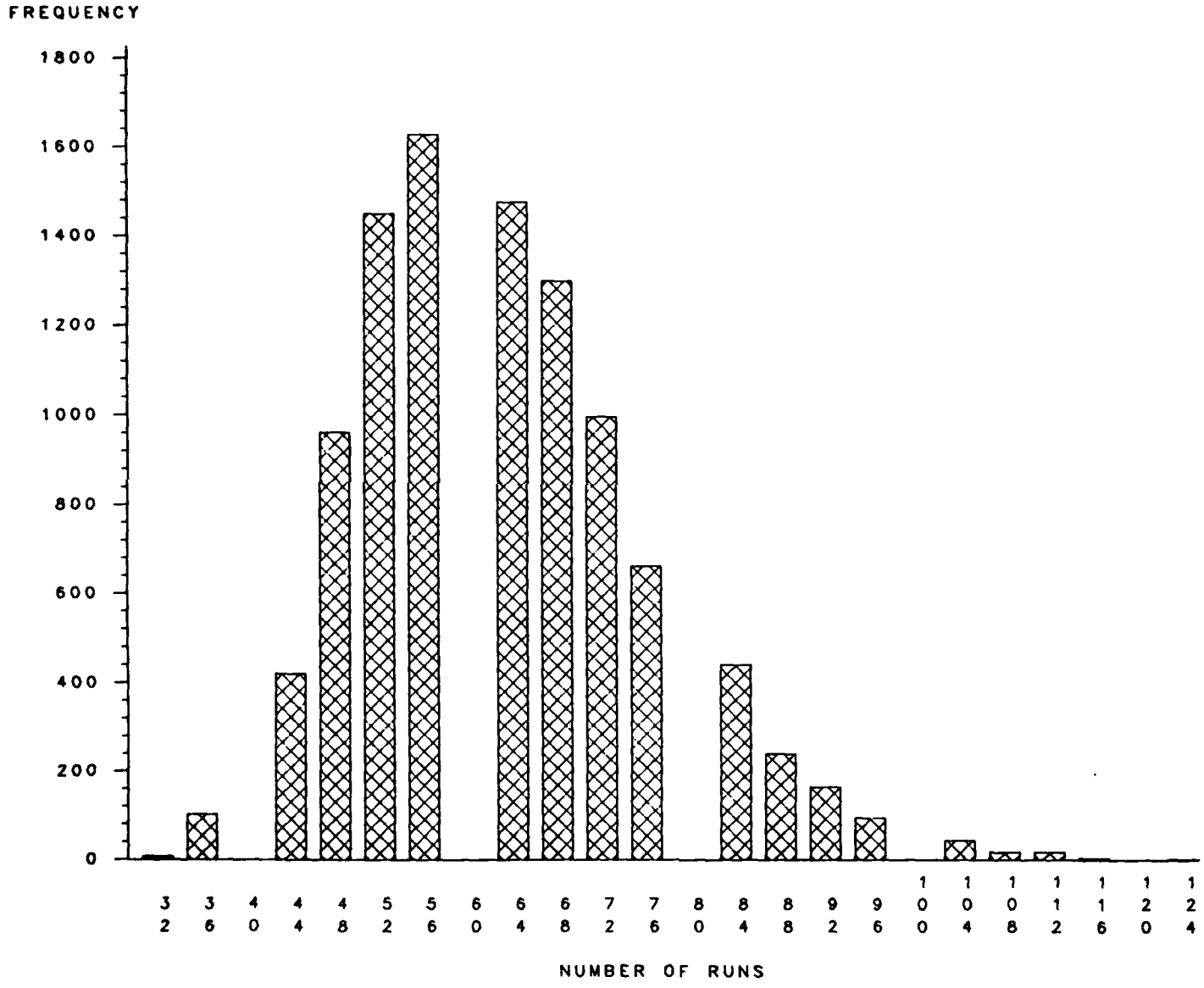


Figure 4. Histogram of Number of Runs of GS(5, 0.002, 0.1667) Strategy ($\bar{N}$=62.9, S_N=12.9).

APPENDIX

The random variables $e_1, e_2, \ldots, e_r$ are said to be exchangeably distributed if the joint distribution of $e_{\pi 1}, e_{\pi 2}, \ldots, e_{\pi r}$ is the same as the joint distribution of $e_1, e_2, \ldots, e_r$ for all permutations π of $(1, 2, \ldots, r)$.

In the ordinary linear model (OLM) we assume that the error terms are i.i.d. normal random variables. In the exchangeable linear model (ELM) we assume exchangeably normally distributed errors. Following Arnold (1981; pp. 232-238), the ELM is equivalently the model in which we observe $\underline{Y} \sim N_r(\underline{\mu}, \sigma^2 \underline{A}(\rho))$, where $\underline{\mu}$ is an rxl mean vector and $\underline{A}(\rho)$ has the following form

$$\underline{A}(\rho) = \begin{bmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \rho & \rho & \ldots & 1 \end{bmatrix} .$$

A key result derived by Arnold (1981) is that in an ELM one-way analysis of variance, equality of level means can be validly tested with the same F-tests customarily used in the OLM. As an aside, we note that the ELM is simply a repeated measures model with only one individual.

We now proceed to show that the RB model when analyzed with separate F-tests has the same covariance structure as the ELM. In matrix notation, model (2.1) can be written compactly as $\underline{y} = \beta_0 \underline{1} + \underline{X\beta} + \underline{\varepsilon}$ where $\underline{1}$ is an Nxl vector of +1's and $\underline{\varepsilon}$ is an Nxl vector $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$ of i.i.d. $N(0, \sigma_\varepsilon^2)$ error terms. We wish to test the hypothesis $H_0: \beta_j = 0$ versus

$H_1: \beta_j \neq 0$ with a simple F-test (or, equivalently, a two-sample t-test) applied to the observations at the high (+1) and low (-1) levels of the $j^{\underline{th}}$ factor. Without loss of generality, we assume that the observations are indexed so that $\{y_i;\ i \leq N/2\}$ have $x_{ij} = +1$ and $\{y_i;\ i \geq N/2\}$ have $x_{ij} = -1$.

Thus, for $i \leq N/2$ we have $y_i = \beta_0 + \beta_j + e_i$, and for $i > N/2$ we have $y_i = \beta_0 - \beta_j + e_i$ where $e_i = \sum_{\substack{m=1 \\ m \neq j}}^{K} \beta_m x_{im} + \varepsilon_i$. It can readily be shown that $\underline{y} = (y_1, y_2, \ldots, y_N)'$ has mean vector $\underline{\mu}$ and variance-covariance matrix $\underline{\Sigma}$ given by

$$\underline{\mu} = (\beta_0 + \beta_j, \ldots, \beta_0 + \beta_j, \beta_0 - \beta_j, \ldots, \beta_0 - \beta_j)',$$

and

$$\underline{\Sigma} = (\tau_j^2 + \sigma_\varepsilon^2)\underline{A}(\rho),$$

where $\tau_j^2 = \sum_{\substack{m=1 \\ m \neq j}}^{K} \beta_m^2$ and $\rho = -\tau_j^2 / \left[(N-1)(\tau_j^2 + \sigma_\varepsilon^2)\right]$.

We see, then, that the RB model has the same covariance structure as the ELM defined earlier, setting $\sigma^2 = \tau_j^2 + \sigma_\varepsilon^2$. The only difference between the two models is that the errors (e_i) in the RB model are not precisely joint normal. We suspect, however, that this violation has little effect on the F-test for two reasons: (1) Arnold (1983) has demonstrated asymptotic validity against nonnormality for tests of this type for the repeated measures model, of which the ELM is a special case. (2) Nonnormality generally has a small effect on tests about means in the presence of balanced sampling, zero skewness, and zero kurtosis. In the RB model, each e_i has zero skewness and kurtosis given by $-2 \sum_{\substack{m=1 \\ m \neq j}}^{K} \beta_m^4 / (\tau_j^2 + \sigma_\varepsilon^2)^2$ which is clearly dominated by the term in the

denominator.

Using results derived by Arnold (1981) for the ELM, it can be shown that the appropriate noncentrality parameter for our testing problem is $\delta = N\beta_j^2/\left[(\tau_j^2 + \sigma_\varepsilon^2)(1-\rho)\right]$, where ρ is as defined previously. Accordingly, an approximation to the power of an RB(N,α_r) strategy for detecting the $j^{\underline{th}}$ factor is given by the expression:

$$\text{Power} = P\{F^* \geq F(1-\alpha_r;1,N-2)\}$$

where F^* has a noncentral F-distribution with (1,N-2) degrees of freedom and noncentrality parameter δ.

REFERENCES

Anscombe, F. J. (1959). "Quick Analysis Methods for Random Balance Screening Experiments." Technometrics, 1, 195-209.

Anscombe, F. J. (1963). "Bayesian Inference Concerning Many Parameters with Reference to Supersaturated Designs." Bulletin of The International Statistical Institute, 40, 721-733.

Arnold, S. F. (1979). "Linear Models with Exchangeably Distributed Errors." Journal of the American Statistical Association, 74, 194-199.

Arnold, S. F. (1981). The Theory of Linear Models and Multivariate Analysis. New York: John Wiley & Sons, Inc.

Arnold, S. F. (1983). "The Asymptotic Validity of Invariant Procedures for the Repeated Measures Model and Multivariate Linear Model." To Appear in Journal of Multivariate Statistics.

Booth, K. H. V. and Cox, D. R. (1962). "Some Systematic Supersaturated Designs." Technometrics, 4, 489-495.

Box, G. E. P. (1959). "Discussion of the Papers of Messrs. Satterthwaite and Budne." Technometrics, 1, 174-180.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). Statistics for Experimenters. New York: John Wiley & Sons, Inc.

Budne, T. A. (1959a). "The Application of Random Balance Designs." Technometrics, 1, 139-155.

Budne, T. A. (1959b). "Random Balance: Part I -- The Missing Statistical Link in Fact Finding Techniques." Industrial Quality Control, 15, 5-10.

Kleijnen, J. P. C. (1975). "Screening Designs for Poly-Factor Experimentation." Technometrics, 17, 487-493.

Li, C. H. (1962). "A Sequential Method for Screening Experimental Variables." Journal of the American Statistical Association, 57, 455-477.

Mauro, C. A. (1983a). "On the Performance of Two-Stage Group Screening Experiments." Desmatics, Inc. Technical Report No. 113-9. To appear in Technometrics.

Mauro, C. A. (1983b). "A Comparison of Random Balance and Two-Stage Group Screening Designs - Part I." Desmatics, Inc. Technical Report No. 113-12.

Mauro, C. A. and Smith, D. E. (1982). "The Performance of Two-Stage Group Screening in Factor Screening Experiments." Technometrics, 24, 325-330.

Mauro, C. A. and Smith, D. E. (1984). "Factor Screening in Simulation: Evaluation of Two Strategies Based on Random Balance Sampling." Management Science, 30, 209-221.

Plackett, R. L. and Burman, J. P. (1946). "The Design of Optimum Multifactor Experiments." Biometrika, 33, 305-325.

Patel, M. S. (1962). "Group Screening with More than Two Stages." Technometrics, 4, 209-217.

Satterthwaite, F. W. (1959). "Random Balance Experimentation." Technometrics, 1, 111-137.

Smith, D. E. and Mauro, C. A. (1984). "Design of Experiments in Simulation." Desmatics, Inc. Technical Report No. 113-13. To Appear in American Journal of Mathematical and Management Sciences.

Watson, G. S. (1961). "A Study of the Group Screening Method." Technometrics, 3, 371-388.

Youden, W. J., Kempthorne, O., Tukey, J. W., Box, G. E. P., and Hunter, J. S. (1959). "Discussion of the Papers of Messrs. Satterthwaite and Budne," Technometrics, 1, 157-193.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 113-14	2. GOVT ACCESSION NO. AD-A137 994	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A COMPARISON OF RANDOM BALANCE AND TWO-STAGE GROUP SCREENING DESIGNS - PART II		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Carl A. Mauro & Kevin C. Burns		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0650
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P.O. Box 618 State College, PA 16804		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-467
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE February 1984 13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Factor Screening
Random Balance Designs
Two-Stage Group Screening Designs
Supersaturated Experiments

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we focus on the problem of supersaturated (fewer runs than factors) screening experiments. We consider two major types of designs which have been proposed in this situation, namely, random balance and two-stage group screening. We discuss the relative merits and demerits of each strategy. In addition, we compare the performance of these strategies by means of a case study in which K=100 factors are screened in N=20,42,62, and 84 runs.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)